MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

SP 21

Neyman's model of competing risks applied to a biological problem.

by

J. de Boer and Prof. Dr J. Hemelrijk.

Contents

1. Summary and introduction.

Limnaea-eggs in two different stages of development, A and B, are treated with Lithium (Li) at temperatures of 30°, 25°, 20° and 15° C. Under the influence of the Lithium the eggs, when nursed, sometimes show an abnormality in their development called exogastrulae and sometimes headaberrations. The Li may also influence the death rate. An investigation was carried out to examine the effect of a change of temperature on the influence of the Li [1]). To this end the number of eggs of egg-masses with exogastrulae, eggs with headaberrations, dead and normal eggs, was observed at a number of different moments after the Li was applied.

To answer the question whether the influence of the Li, especially on the number of eggs with exogastrulae or with headaberrations, is affected by the temperature or not; the results of this experiment have been analysed statistically by means of the NEYMAN-scheme of competing risks. A method like this is necessary, because the same egg cannot exhibit exogastrulae and headaberrations. Also an egg that dies, might have shown exogastrulae or headaberrations, had it lived longer.

We find that the "net"rates of the risk of headaberrations and of the risk of exogastrulae in stages A and B respectively, increase and decrease respectively with rising temperature.

2. Crude and net rates of risk.

In the terminology of NEYMAN which we shall adopt here, the eggs are, after treatment with Li, exposed to four differen so-called "risks", namely the risk of staying normal, dying and exhibiting exogastrulae or headaberrations.

In both stages A and B of development, separately, 21 egg-masses have been observed, each consisting of 4 × 13 eggs; 13 eg are treated at each of 4 different temperatures. The stages A and B are investigated separately by identical methods.

If we count the number of eggs of an egg-mass, which at th end of the experiment are in one of the four possible states (where we count a death following an exogastrula or headaberration as an exogastrula or headaberration respectively), and divide this number by the total number of eggs in the egg-mass (i.e. by 13), we get an estimate of the so-called "crude" rate of the risk corresponding to that state, i.e. the probability of succumbing to that particular risk when a normal egg is exp ed to a number of competing risks. We are, however, not interested in the crude rates only. Our main interest is in the net rate of a risk, defined as the probability in st te of normality, of succumbing to that risk when all other risks, except staying normal, are eliminated. A mathematical specification of this concept will be given later.

As we are interested in detecting differences in the net rates of risk at different temperatures, and as only the crude rates can be estimated directly from the observations, we have to establish relations between the crude and net rates. To do this, we follow the method of NEYMAN [1] 2). These relations then enable us to draw conclusions about the net rates from th results about the crude rates which follow from a routine analysis of the observations.

3. Mathematical model.

This section and the first part of section 4 are concern ed with considerations, which are valid for every egg-mass sep ately. Thus we consider one egg-mass at a time. For every egg-mass the last exogastrula occurred before the first headaberration. The time after treatment with Li may therefore be divi ed in three periods (a, b and c), exogastrulae and headaberra tions respectively occurring only in the first and third period respectively, and the second one containing only cases of deat Death may also occur in the other periods. When speaking of th time of an exogastrula, death or headaberration, we mean the first moment when this new state has been observed.

According to NEYMAN each of the three periods is divided into a large number of small intervals of time, each of length Δt . During such a time Δt a normal egg may remain normal or may pass into one of the three other states. Transition from any of these three states to any other state is impossible, deaths after exogastrulae or headaberrations not being counted as such (cf. § 2). We denote the probability of the possible

---

2) Numbers within brackets refer to the references at the end of this paper.

transitions during Δt by q_{ik}, i indicating the period (i = a, b or c) and k the state of the egg at the end of the interval Δt considered, with

k = 0 corresponding to normality (N)
k = 1 " " exogastrulae (E)
k = 2 " " headaberrations (H)
k = 3 " " death (D)

The transition probabilities are taken to be constant during each of the three periods, but possibly different for different periods.

Considering each of the three periods, as well as a, b and c together, as a scheme of competing risks, we have crude rates of risk k (k = 0,1,2,3) in a, b and c respectively (denoted by Q_{ak}, Q_{bk} and Q_{ck}), as well as in the periods together (denoted by Q_k).

We now have the following scheme of possible transitions:

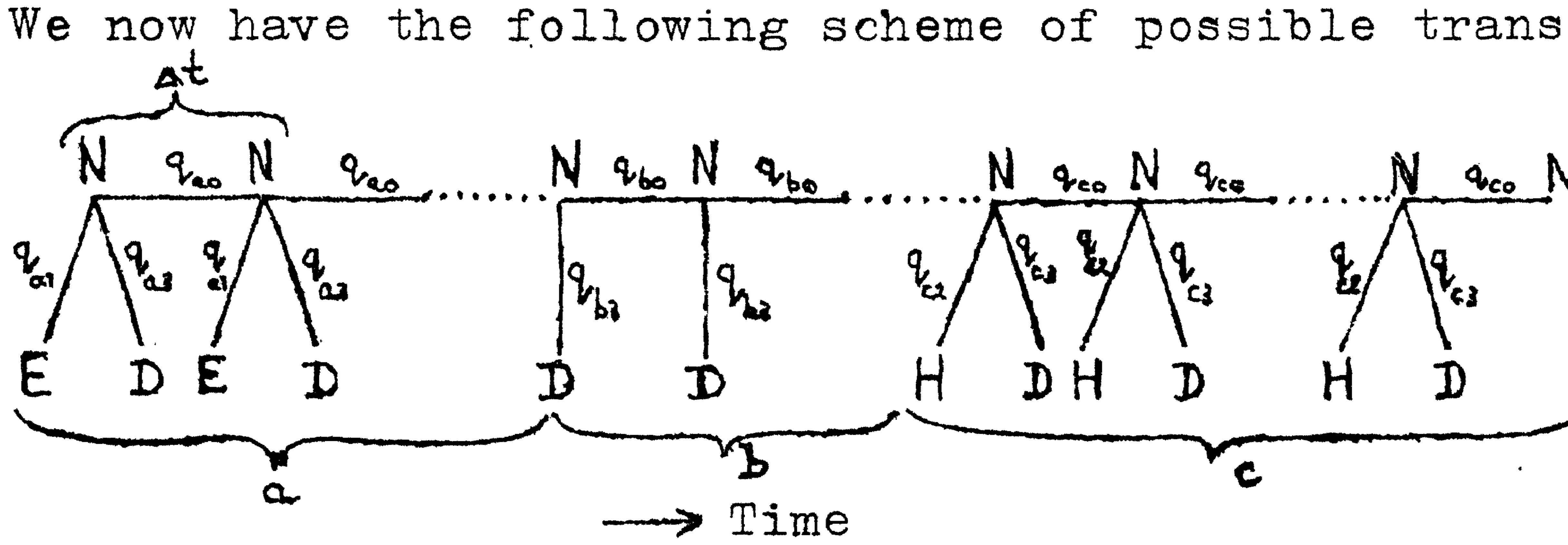

Fig. 1

Possible transitions of a Limnaea-egg after treatment with Li.

It is clear that this model is a simplification of reality as every mathematical model always is. It is easy to build more complicated models, but difficult to handle them and therefore we have contended ourselves with this one.

4. <u>Relations between net and crude rates of risks.</u>

We divide each period in a number (say M) of parts, each of length Δt. It is assumed that the probabilities of transfer q_{ik} are proportional (at least in the limit) with the length of the time interval Δt considered. This means that the quantities $\lambda'_{ik} \stackrel{def.}{=} q_{ik} M$ (k = 1,2,3) tend to a limit λ_{ik} if M tends to infinity:

$$\lim q_{ik} M = \lambda_{ik} \qquad (k = 1,2,3)$$

(Here and in what follows, all limits are for $M \to \infty$)

Now we are in a position to express the crude rates Q_{ik} in terms of λ_{ik}. The rates Q_0 and Q_3 of normality and death respectively are not very interesting from a biological point of view and may be left out of the investigation. Furthermore, exogastrulae only occurring in the first period and headaberrations only in the third one, we have $Q_{a1} = Q_1$ and $Q_{c2} = Q_2$. Thus we only need to compute the crude rates for the periods separately.

Considering the M time intervals in period a, we get:

$$(1)\quad Q_{a0} = \lim q_{v}^{M} = \lim (1 - q_{a1} - q_{a3})^{M} = \lim \left(1 - \frac{\lambda'_{a1} + \lambda'_{a3}}{M}\right)^{M} = e^{-\lambda_{a1} - \lambda_{a3}}.$$

is the limit of the sum of the probabilities of transfer from state 0 to state 1 in the 1st, 2nd, ..., Mth element of time of period a. Thus:

$$(2)\quad Q_{a1} = \lim \sum_{k=1}^{M} q_{a0}^{k-1} q_{a1} = \lim q_{a1} \frac{1 - q_{a0}^{M}}{1 - q_{a0}} = \lim \frac{q_{a1}}{q_{a1} + q_{a3}} (1 - q_{a0}^{M}) = \frac{\lambda_{a1}}{\lambda_{a1} + \lambda_{a3}} (1 - Q_{a0}).$$

Considering period b separately as a scheme of competing risks, we obtain:

$$(3)\quad Q_{b0} = \lim q_{b0}^{M} = \lim (1 - q_{b3})^{M} = \lim \left(1 - \frac{\lambda'_{b3}}{M}\right)^{M} = e^{-\lambda_{b3}}.$$

For period c we get in the same way:

$$(4)\quad Q_{c0} = e^{-\lambda_{c2} - \lambda_{c3}},$$

$$(5)\quad Q_{c2} = \frac{\lambda_{c2}}{\lambda_{c2} + \lambda_{c3}} (1 - Q_{c0}).$$

We now introduce the concept of net rate of risk mathematically. The net rate of risk k (k = 1,2,3) in period i (i=a,b c) denoted by P_{ik}, is defined as the probability of transition, in period i, from state 0 to state k, when all other risks h (h = 1,2,3; h≠k) are eliminated, the transition probabilities of these other risks being added to the transition probability of staying normal. Or, in other words, P_{ik} is the crude rate when q_{i0} is replaced by

$$q_{i0}^{*} = 1 - q_{ik} \qquad (i = a, b, c).$$

The net rates for the three periods taken together will be denoted by P_k (k = 1,2,3). The definition of these net rates is the same as for P_{ik}.

Thus, $1 - P_{ik}$ is the probability, in period i, of staying normal with $q_{i0}^{*} = 1 - q_{ik}$ as the transition probability of staying normal. Therefore we have

$$(6)\quad 1 - P_{ik} = \lim (1 - q_{ik})^{M} = \lim \left(1 - \frac{\lambda'_{ik}}{M}\right)^{M} = e^{-\lambda_{ik}}.$$

We use this formula to compute $P_{a1}(=P_1)$ and $P_{c2}(=P_2)$. From (2) we get

(7) $\lambda_{a1} = \frac{Q_{a1}}{1-Q_{a0}}(\lambda_{a1}+\lambda_{a3})$.

Thus, with (6), (2) and (1):

(8) $P_{a1} = 1-e^{-\lambda_{a1}} = 1-e^{-(\lambda_{a1}+\lambda_{a3})\frac{Q_{a1}}{1-Q_{a0}}} = 1-Q_{a0}^{\frac{Q_{a1}}{1-Q_{a0}}}$.

In the same way, with (4), (5) and (6):

(9) $P_{c2} = 1-Q_{c0}^{\frac{Q_{c2}}{1-Q_{c0}}}$.

All this has been derived for one egg-mass. In different egg-masses the length of the periods a, b and c may be different [3], but this does not affect the definitions of crude and net rates of risks, nor the derivation of the formulas (7), (8) and (9). In fact, all rates of risk for one egg-mass refer to the intervals of that particular egg-mass, and may differ from the corresponding rates of risk of other egg-masses. Thus, if we want to test the hypothesis, that the net rate P_1 of exogastrulation (or P_2 of headaberrations) is not influenced by the difference of the four temperatures used in the experiment, the right thing to do is to test this hypothesis for every egg-mass separately and to combine the results of these tests into one overall test. This is easy, because the experiments with the 21 egg-masses are completely independent.

Under the hypothesis H_0 that P_1 is not influenced by the temperature, we have $P_1 = P_{a1} = C$, where C is the same for the four temperatures (though possibly different for different egg-masses). Hence, from (8) we find, for every egg-mass separately:

(10) $Q_{a1} = \frac{\ln(1-C)}{\ln Q_{a0}}(1-Q_{a0})$.

An anlogous equation is valid for Q_{c2} and Q_{c0}.

By differentiation we find that for every C, Q_{a1} is a monotonic increasing function of Q_{a0} (cf. fig. 2).

---

3) In fact these periods differed very little for the 21 egg-masses and the periods could roughly be taken equal for all of them, except only a few cases.

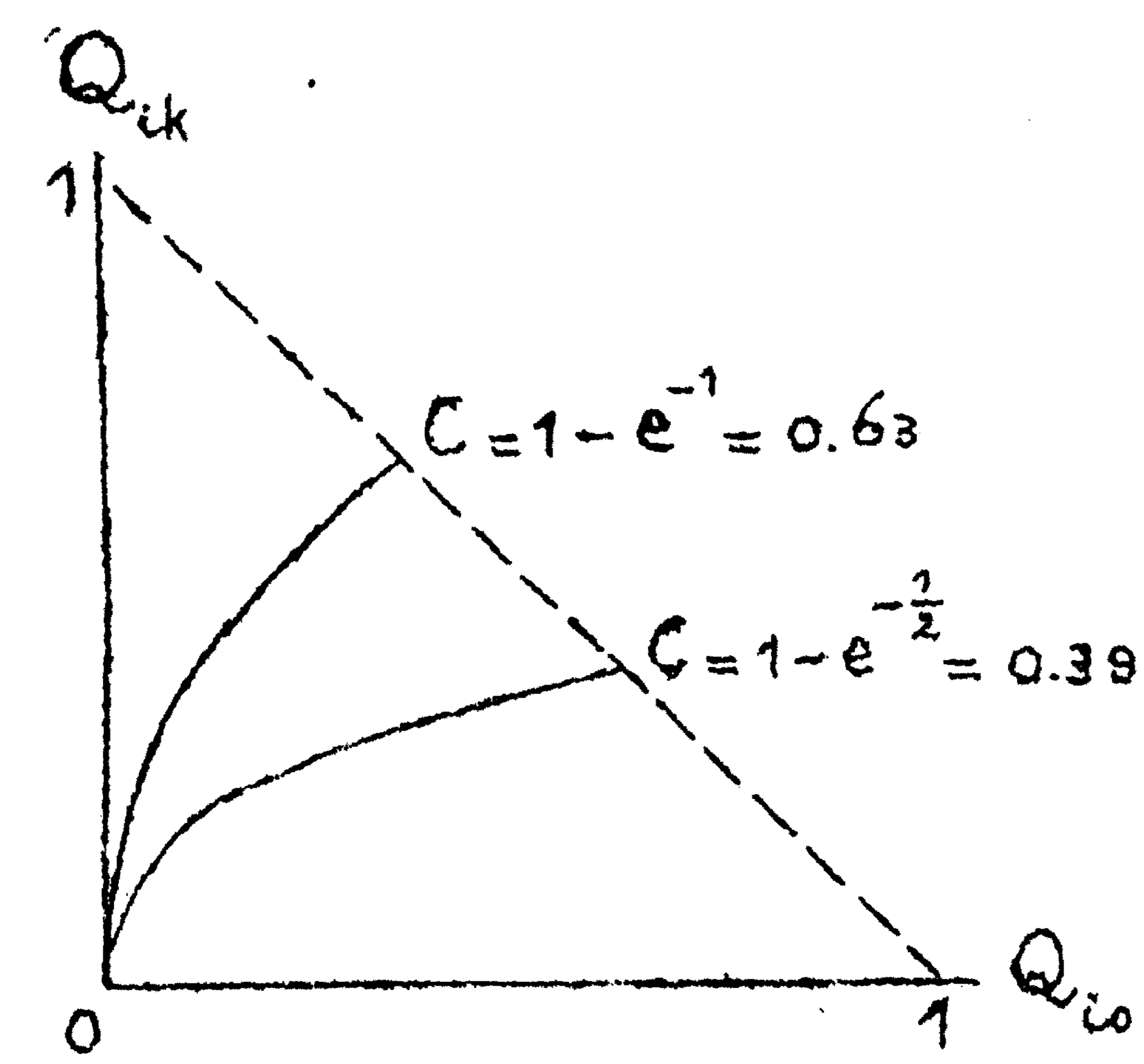

Fig. 2

Graphical representation of the function

$$Q_{ik} = \frac{\ln(1-C)}{\ln Q_{io}} (1 - Q_{io}).$$

It follows that H_0 can be rejected if this monotony is refuted by the observations. Moreover we remark, that for any two values of C between 0 and 1, the curve with the highest value of C lies above the other curve.

An analogous argument holds for the hypothesis that P_2 is not affected by the temperature.

5. Application.

The result of the experiments described in section 1 may be found in table I. This table contains both the results for stage A and stage B, which will be investigated separately.

Table 1

Number of exogastrulae (E), headaberrations (B) and deaths (D) in the three periods a, b and c for each of 21 egg-masses after treatment with Lithium at four temperatures of 15°, 20°, 25° and 30° C.

Stage	No. of egg-mass	a								b				c							
		Number of E				Number of D				Number of D				Number of D				Number of H			
		15°	20°	25°	30°	15°	20°	25°	30°	15°	20°	25°	30°	15°	20°	25°	30°	15°	20°	25°	30°
A	1			2	7								1			1			1	2	1
	2				1				2			2		1					1	1	1
	3				1	1		2	1			1	1	1	1			1		1	3
	4				2															1	2
	5																1	1			
	6				1	1			1												
	7																				
	8	1	8	5	4		1		1					2	1	6				1	
	9							1													
	10								1						2						
	11			1				1				1						1	1	5	3
	12			3	6				1			1	2			1	1			3	
	13			2	1	1	1														
	14						2														
	15					1			1												
	16			1	1			1	1			1							2	1	
	17				1		1								1						
	18			1											1	1	1			2	
	19						1	1													
	20		2	2						1				1					1		
	21		1		4									1		2			1	1	3
B	1	1	2												1			1			
	2	6	1							1	1			1				2			
	3	2	1												1	1					1
	4	4	4		1					1	1			3	2	1		2		1	2
	5	6	3	2	2	1									1				2	4	1
	6							2													
	7	2	1								1				2				4		
	8	1										1	1					1			
	9	11	8	3			1		1			1							1	2	1
	10	6	8	7	2												1	3	1	2	
	11					1	2											1	1		1
	12	3	2					1							1			5	3		
	13										3				4			1		2	3
	14	3	5												4				2	2	2
	15					1												2		2	2
	16	3		1	1										2			2		1	1
	17	1	1	1										1			1	5			
	18	2	1				3			3				3		1	1	3	1	3	1
	19	1	2				1							2	2	2	1	1	2	3	1
	20	2							1	1				1	1	1	1	1	3	5	1
	21	1													3		1	2			

We now test for stage A the hypothesis that the crude rate is invariant under a change of temperature by means of FRIEDMAN's method of m rankings (cf. e.g. [2]), i.e. we set up a scheme with 21 rows (the egg-masses) and 4 columns (the temperatures), putting in plot (i,k) the number of eggs staying normal during period A in egg-mass i at temperature k. These numbers are then ranked according to their size in each row separately and the ranks are added together for each column. Rows, where all eggs remained normal were omitted as useless. From the column totals, obtained in this way, a statistic measuring the concordance between rows is computed, which, under the hypothesis tested, is approximately distributed as χ^2 with a known number of degrees of freedom. With large values of this coefficient as critical values the onesided tail error P has been found from a table of the χ^2-distribution.

The same method was applied to test the hypothesis of invariance for Q_{ao} in stage B, and for Q_{a1} in both stages. In the latter case we put in plot (i k) the number of eggs in egg-mass i exhibiting exogastrulae at temperature k. The results are summed up in table II.

Table II

Tail probabilities in testing the invariance of Q_{ao} and Q_{a1} at different temperatures.

Stage	Crude rate	Column totals				P
		15°	20°	25°	30°	
A	Q_{ao}	59	51	42,5	37,5	0,010↓
	Q_{a1}	24	30,5	39,5	44	0,003↑
B	Q_{ao}	30,5	41	61,5	67	$<10^{-4}$ ↑
	Q_{a1}	61,5	50,5	31	27	$<10^{-4}$ ↓

An arrow, pointing upwards or downwards respectively, means that the concerning crude rate increases or decreases respectively with increasing temperature.

All tail probabilities being small, the hypothesis of invariance under changes of temperature must be rejected for all cases. It further follows from the column totals in table II, that Q_{ao} decreases with increasing temperature in stage A, but increases in stage B. Q_{a1} increases with temperature in stage A and decreases in stage B. By means of (10) and fig. 2 we may thus conclude, that C is not invariant either for changes in temperature but that C decreases with increasing temperature in stage A and increases in stage B. This means, that the net rate of risk of exogastrulation $P_1 = P_{a1} = 1 - C$ increases with increasing temperature in stage A and decreases in stage B.

The above procedure fails for Q_{c0} and Q_{c2} because of the unequal sizes of the egg-masses at the beginning of period c. Therefore the hypothesis of invariance is tested in another way, by means of 2x2-tables for every egg-mass separately, the results being combined afterwards.

Let us take e.g. the test for invariance of Q_{c0} under changes of temperature. For every egg-mass we construct two 2x2-tables, with the temperatures 15^o and 25^o (and 20^o and 30^o respectively) as one dichotomy and the results "normal" or "not normal at the end of period c" as the other dichotomy. For egg-mass no 1 for instance and for the temperatures 15^o and 25^o, the number of eggs at the beginning of period c may be seen from table I to be 13 and 10(= 13-2-1) respectively. Of these eggs 13 and 7(= 10-1-2) are normal at the end of period c. Thus the 2X2-table for this case is:

	normal	not normal	total
15^o	13	0	13
25^o	7	3	10
total	20	3	23

It is a well known fact, that e.g. the number in the left hand top corner of this table (i.e. in this case the number of normal eggs at 15^o C) has, under the hypothesis of independence tested, a hypergeometric distribution with known mean and variance. Denoting this number by $\underline{x}$ in each of the 2X2-tables pertaining to the investigation of the invariance of Q_{c0} under changes of temperature, we add all values of $\underline{x}$, denoting the sum by $\underline{s}$, adding also their means and variances. These last two sums then give us the mean $\mathcal{E}\underline{s}$ and variance $\sigma^2(\underline{s})$ of $\underline{s}$ under the hypothesis tested and using the approximately normal distribution of $\underline{s}$, we may test the hypothesis of invariance of by means of the statistic

$$\frac{\underline{s} - \mathcal{E}\underline{s}}{\sigma(\underline{s})}$$

which is approximately normally distributed with zero mean and unit variance if the hypothesis tested is true.

This procedure was applied to Q_{co} and Q_{c2} in the stages A and B of development separately. The summation of the values of x was effected for the pairs of temperatures (15°, 25°) and (20°, 30°) separately, and for all 2×2-tables pertaining to one experiment together. The results, with the onesided tail probability P in the last column, may be found in table III. It may be remarked, that this method could also be applied to period a instead of the method of m rankings. In fact for period a still other methods are possible. The method of m rankings was chosen because of its simplicity and proved to be adequate. Had it been possible to apply it to period c too, then we would certainly have done so.

Table III

Results of testing the invariance of Q_{co} and Q_{c2} at different temperatures.

Stage	Crude rate	Compaired temperatures	s	$\mathcal{E}_s$	$\sigma(s)$	P
A	Q_{co}	15°- 25°	102	114,146	2,750	10^{-5} ↓
		20°- 30°	140	142,403	2,592	0,177 ↓
		combined	242	256,549	3,779	10^{-4} ↓
	Q_{c2}	15°- 25°	18	9,415	2,169	10^{-4} ↑
		20°- 30°	13	8,820	2,104	0,023 ↑
		combined	31	18,235	3,022	10^{-5} ↑
B	Q_{co}	15°- 25°	197	187,015	3,739	0,004 ↑
		20°- 30°	203	178,885	3,659	10^{-9} ↑
		combined	400	365,900	5,232	10^{-9} ↑
	Q_{c2}	15°- 25°	27	33,442	3,517	0,004 ↓
		20°- 30°	17	20,804	2,858	0,092 ↓
		combined	44	54,246	4,532	0,012 ↓

The results in table III clearly indicate, that the hypothesis tested of invariance of the crude rates should be rejected immediately. In fact we may safely draw the conclusion, that in stage A Q_{co} decreases and Q_{c2} increases with rising temperature, while in stage B the situation is reserved. According to fig. 2 this means, that the net rate of risk of headaberrations (P_2) increases with rising temperature in stage A and decreases in stage B.

6. Conclusions.

Distinguishing between the crude and net rate of a risk, and having established relations between these two kinds of rates, we have examined the invariance of net rates by testing the invariance of crude rates. Finding an opposite trend in the crude rates with rising temperature, we conclude, in connection with the above mentioned relations, that the corresponding net rate cannot be invariant under varying temperature.

In this way we found that the net rates of the risk of headaberrations and of exogastrulation increases in stage A and decreases in stage B when the temperature rises.

References.

[1] J. NEYMAN, First course in probability and statistics, Henry Holt and Co., New York, 1950.

[2] M.G. KENDALL, Rank correlation methods, Charles Griffin and Co. Ltd., London, 1948.